Université de Montpellier Année 2023-2024
HAX912X - Modèles linéaires généralisés

# Feuille de Travaux Dirigés 2

# Modèles linéaires généralisées

**Exercice 1**

On considère $n$ variables aléatoires indépendantes $y_1, \ldots, y_n$ telles que $y_i \sim \mathcal{N}\left(\exp(\beta_0 + \beta_1 x_i), \sigma^2\right)$.

**1** Montrer qu'il s'agit d'un modèle linéaire généralisée. En suivant les notations du cours quant aux familles exponentielles à un paramètre de nuisance, on explicitera les paramètres $\phi$, $\theta$, les fonctions $\gamma$ et $b$ ainsi que les régresseurs à considérer.

**2** Montrer que ce n'est pas le lien canonique qui a été choisi.

**3** On suppose dans la suite que $\sigma^2 = 1$, donner la log-vraisemblance et les équations de vraisemblance. Pouvons-nous calculer les expressions analytique des estimateurs du maximum de vraisemblance de $\beta_0$ et $\beta_1$ ?

**Exercice 2**

On considère $n$ variables aléatoires indépendantes de loi de Poisson de paramètre $\mu_i = \mathbb{E}(y_i)$. Nous supposons que $\log(\mu_i) = \beta_0 + \beta_1 x_i$ où $\beta_0$ et $\beta_1$ sont des paramètres inconnus.

**1** Montrer qu'il s'agit d'un modèle linéaire généralisée.
Est-ce le lien canonique qui a été choisi ?

**2** Donner la log-vraisemblance et les équations de vraisemblance. Pouvons-nous calculer les expressions analytique des estimateurs du maximum de vraisemblance de $\beta_0$ et $\beta_1$ ?

**Exercice 3**

On considère $n$ variables aléatoires indépendantes $y_1, \ldots, y_n$ telles que $y_i$ est distribué suivant une loi négative binomiale de paramètres $(h, \pi_i)$. Nous supposons que $h$ est fixé. La loi négative binomiale modélise, dans le contexte d'une suite d'épreuves de Bernoulli indépendantes, le nombre d'essais nécessaires pour obtenir $h$ succès, $\pi_i$ représente la probabilité de succès. Nous avons

$$f(y_i; \pi_i) = C_{y_i-1}^{h-1} \pi_i^h (1-\pi_i)^{y_i-h} \mathbf{1}_{\{h,h+1,\ldots\}}(y_i)\,.$$

Pour tout $i = 1, \ldots, n$, on suppose que $\log(\pi_i/(1-\pi_i)) = \beta_0 + \beta_1 x_i$ où $x_i \in \mathbb{R}$ est supposé connu, $\beta_0$ et $\beta_1$ sont des paramètres inconnus.

**1** Montrer qu'il s'agit d'un modèle linéaire généralisée. La fonction de lien canonique a-t-elle été utilisée ?

**2** Donner la log-vraisemblance et les équations de vraisemblance. Pouvons-nous calculer les expressions analytique des estimateurs du maximum de vraisemblance de $\beta_0$ et $\beta_1$ ?

**3** Modifier le modèle en utilisant le lien canonique et donner les équations de vraisemblance.

**Exercice 4**

On considère un $n$-échantillons $(y_1, x_1), \ldots, (y_n, x_n)$ du couple $(y, x)$ où $y$ est une variable aléatoire à valeurs dans $\{0, 1\}$ et $x$ est une variable fixée prenant également les valeurs $\{0, 1\}$. On suppose qu'il existe une variable latente $y_i^*$ telle que $y_i^*|x_i \sim \mathcal{N}(\beta_0 + \beta_1 x_i, 1)$ et

$$y_i = \begin{cases} 1 & \text{si} \quad y_i^* \leq 0 \\ 0 & \text{si} \quad y_i^* > 0 \end{cases} .$$

On note $\Phi(\cdot)$ la fonction de répartition de la loi normale centrée réduite ($\Phi(u) = \mathbb{P}(\mathcal{N}(0,1) \leq u)$).

**1** Montrer qu'il s'agit d'un modèle linéaire généralisée. En suivant les notations du cours quant aux familles exponentielles à un paramètre de nuisance, on explicitera les paramètres $\phi$, $\theta$, les fonctions $b$ et $c$ ainsi que la fonction $\gamma$ et les régresseurs à considérer.

**2** Est-ce le lien canonique qui a été choisi ?

**3** Nous disposons de 100 observations $(y_i, x_i)$ décrites dans la table de contingence suivante

| | $y = 0$ | $y = 1$ |
|---|---|---|
| $x = 0$ | 10 | 30 |
| $x = 1$ | 20 | 40 |

On note $\alpha_0 = \Phi(\beta_0)$ et $\alpha_1 = \Phi(\beta_0 + \beta_1)$ donner la log-vraisemblance en fonction de $\beta_0$ et $\beta_1$.

**4** Calculer l'estimateur du maximum de vraisemblance du couple $(\alpha_0, \alpha_1)$.